



Audio Engineering Society

Convention e-Brief 651

Presented at the 151st Convention
2021 October, Online

This Engineering Brief was selected on the basis of a submitted synopsis. The author is solely responsible for its presentation, and the AES takes no responsibility for its contents. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Audio Engineering Society.

Production Tools for the MPEG-H Audio System

Philipp Eibl¹, Yannik Grewe¹, Daniela Rieger¹, and Ulli Scuda¹

¹ *Fraunhofer Institute for Integrated Circuits IIS, Am Wolfsmantel 33, 91058 Erlangen, Germany*

Correspondence should be addressed to Philipp Eibl (philipp.eibl@iis.fraunhofer.de)

ABSTRACT

Next Generation Audio (NGA) systems like MPEG-H Audio rely on metadata to enable a wide variety of features. Information such as channel layouts, the position and properties of audio objects, and user interactivity options are only some of the data that can be used to improve consumer experience. Creating these metadata requires suitable tools, which are used in a process known as "authoring", where interactive features and the options for 3D immersive sound rendering are defined by the content creator. Different types of productions impose specific requirements on these authoring tools, which leads to a number of solutions appearing in the market. Using the example of MPEG-H Audio, this paper will detail some of the latest developments and authoring solutions designed to enable immersive and interactive live and post-productions.

1 Introduction

Next Generation Audio (NGA) systems are developed to provide a personalized and immersive audio experience. To enable end-user personalization and optimized rendering for different kinds of playback device, NGA relies on metadata, which are transmitted with the content and interpreted by the decoder and renderer [1].

Depending on the NGA system, the available metadata feature set varies. This paper covers metadata creation using the example of MPEG-H Audio [2]. The MPEG-H Audio system, based on ISO/IEC 23008-3 [3], supports a wide variety of metadata and functionality for user interactivity. Several tools for metadata creation are available from multiple hardware and software manufacturers. Following the standard ensures interoperability between tools, when using BWF files (BW64 and RF64) with embedded metadata chunks according to ITU-R BS.2088-1[4].

2 Metadata Authoring

With MPEG-H Audio, content creators can offer a variety of options for personalization to their audience. All these advanced audio features can be strictly defined by the content provider during metadata creation in a process called authoring [5, 6].

The authoring process can either take place during content production, e.g., within the audio mixing session, by means of a plug-in for Digital Audio Workstations (DAWs), using hardware-based live authoring tools, or as an independent step.

In file-based post-production scenarios, the result of the authoring process is an intermediate file called MPEG-H Audio Master. It is a multi-channel PCM wav file with associated metadata, either in the form of an Audio Definition Model (ADM)-based [7] Broadcast Wave Format (BWF) file which contains a metadata chunk [4], or in form of an MPEG-H Production Format (MPF) file, where the metadata is encoded into a timecode-like audio signal – the Control Track [8].

The standardization of a serialized version of ADM (S-ADM), based on ITU-R BS.2125-0, for file-based and IP-based transmission in combination with SMPTE 2110, is currently ongoing [9,10].

Authoring during the mixing process enables the elaborate use of metadata for dynamic audio objects that can change position over time, and for other time-line based automation like gain sequences [5].

Authoring of pre-produced content can be done in multiple ways and largely be automated through content server- or cloud-based solutions. This usually requires audio stems to be exported individually, if they are intended to be assigned to different MPEG-H Components to enable interactivity, such as the choice between multiple languages.

Another application of automated enhancement of pre-mixed or legacy content is MPEG-H Dialog+, a DNN-based dialog separation and automatic remixing technology, which was demonstrated in trials with German public broadcasters [11, 12].

3 File-Based Metadata Creation

Metadata used in Next Generation Audio systems offer a myriad of options to describe an audio scene and to specify content and properties of the contained audio signals. Most of them are not mandatory, or are created automatically, for example basic audio scene information or loudness metadata. The desired complexity of the metadata-based functionality depends on the type of content and the production environment.

Since MPEG-H is an open standard, a variety of different tools have emerged in the market, with some striving to enable as many options as possible, while others aim to streamline specific use cases.

Tools for file-based MPEG-H authoring can be classified into different categories, such as stand-alone software, plug-ins, and server- or cloud-based systems.

3.1 Standalone Software and DAW Plug-Ins

There is a variety of tools for creating, editing, and handling of file-based MPEG-H Scenes. They are typically used on local machines. The metadata

creation can either be integrated natively into a host application, such as an audio/video editor [13], or use a dedicated authoring tool, such as [14]. These tools typically export the audio and metadata to an MPEG-H Audio Master in MPF or BWF/ADM formats suitable for archiving, further processing, and encoding.

Other stand-alone tools specialize in specific parts of NGA workflows. An MPEG-H conversion tool, for instance, allows the conversion of various file types into ITU-R BS.2088-1 compliant BWF/ADM files, thus ensuring interoperability between different NGA toolsets and independence from proprietary formats [14]. Other dedicated desktop software includes metadata info tools [15] or encoders.

Plug-ins, on the other hand, seek to enhance the functionality of existing host-software, such as DAWs. Content producers can continue to use their preferred editing and mixing software, which leads to an easier transition from legacy- to NGA-workflows. The additional features introduced by plug-ins do not only enable the authoring of scenes, but also make it possible to monitor downmixes to a variety of formats and preview interactivity and personalization features.

Authoring plug-ins [14, 16, 17] provide many options for metadata-based scene adjustments. These include the creation of different pre-defined audio scene selections called ‘MPEG-H Presets’, multiple language tracks the viewers can choose from, and metadata which allow for a selection based on the playback device’s preference settings, such as a “hearing-impaired” preset.

While personalization improves accessibility, many of these options are mainly angled toward the use in broadcast or video-streaming services.

Recently, plug-ins have been released or announced which aim towards media creation for immersive music, which is increasingly adopted by major music streaming providers [18]. Since current immersive music services offer three-dimensional immersive sound without user interactivity, authoring options enabling interactivity during playback are not required here. Therefore, these tools have opted for a more stream-lined interface optimized for this specific target format. Tools which are conceived with content production for Sony 360 Reality Audio streaming in mind use a

specific subset of MPEG-H Audio, which focusses on audio objects only, instead of on a mixed approach using both channels and objects [19]. For more information on music-focused mixing and authoring plug-ins, see [20, 21].

3.2 Server- and Cloud-Based Solutions

In fast-paced production environments like broadcast stations, the use of server- or cloud-based solutions for metadata generation can help streamline processes. Such services allow for automation of standardized metadata creation workflows. This includes the authoring of MPEG-H interactivity metadata alongside more traditional data, such as loudness information. Broadcasters can choose from automation processes and combine them in a way that suits their typical production requirements [22]. This is especially interesting for recurring shows or program segments, where the required metadata and program components, such as languages or personalization options, follow a consistent layout.

Server- or cloud-based solutions can also offer the encoding of the intermediate production format (e.g., BWF/ADM) into a data-reduced format for distribution via direct integration into a broadcaster's media asset management infrastructure.

4 Authoring in Live Production

Professional broadcast also requires solutions for live production, where metadata – including loudness information – are created and transmitted to an encoder in real-time. Additionally, broadcasters need to be able to monitor audio and metadata rendering at any time.

In these environments, hardware-based solutions in combination with server-based processes are preferred over locally executed software, as they provide higher stability through focus on specific applications. For this purpose, the device class of Audio Monitoring and Authoring Units (AMAUs) has been developed.

These devices allow on-the-fly metadata authoring and metadata passthrough (when playing pre-produced content), creation of loudness metadata,

and modulation of the metadata into a Control Track for real-time transmission to a professional contribution- or emission encoder [23].

Hardware devices which have so far focused on live immersive mixing capabilities for custom layouts, such as planetariums and live shows, can receive MPEG-H capabilities via software update. The addition of authoring, monitoring and delivery functionality can extend the use-cases of such devices into the live-broadcast world [24].

Future hardware can be expected to adapt to specific requirements, e.g., higher signal counts, like the live mixing console recently developed by NHK Science & Technology Research Laboratories [25]. With the support of live transmission of up to 64 audio channels and S-ADM metadata, it also allows monitoring and rendering layouts of up to 22.2, hinting at fewer signal-count restrictions, as future infrastructure moves further towards IP-based workflows.

With ongoing standardization efforts in NGA broadcast and an expanding market for 3D audio, the appeal for hardware- and software manufacturers to develop competitive solutions can be expected to grow [26].

5 Conclusion

With a growing number of production tools in the market and the adoption of ADM as an exchange format between systems, content producers can choose from a variety of tools to optimize their production workflows and take a step from legacy production to working with NGA [27]. Relying on metadata, which are created during the new authoring step, Next Generation Audio systems offer various features, such as personalized and immersive audio. The open standard NGA system MPEG-H Audio allows object-based production workflows for file-based as well as live broadcast environments and is supported in numerous tools such as plug-ins, cloud-based solutions, and hardware-based products. With continued standardization efforts in NGA broadcast and increasing adoption of 3D audio by streaming services, the range of hard- and software solutions can be expected to grow.

References

- [1] Grewe, Y., Simon, C., and Scuda, U. (2018), "Producing Next Generation Audio using the MPEG-H TV Audio System", NAB Broadcast Engineering and Information Technology Conference Proceedings 2018.
- [2] Herre, J. et al. (2014), MPEG-H Audio–The New Standard for Universal Spatial/3D Audio Coding.
- [3] ISO/IEC 23008-3:2019: "Information technology–High efficiency coding and media delivery in heterogeneous environments–Part 3: 3D audio", including ISO/IEC 23008-3:2019/AMD 1:2019 "Audio metadata enhancements" and ISO/IEC 23008-3:2019/AMD 2:2020, "3D Audio baseline profile, corrections and improvements"
- [4] "BS.2088: Long-form file format for the international exchange of audio programme materials with metadata", <https://www.itu.int/rec/R-REC-BS.2088-1-201910-I> Retrieved 6. August 2021.
- [5] Bleidt, R.L. (2017), "Development of the MPEG-H TV Audio System for ATSC 3.0", in IEEE Transactions on Broadcasting, Vol. 63, No. 1
- [6] Scuda, U. et al. (2021), „MPEG-H Audio für Rundfunk, Streaming und Musik“, in bvft Filmtown Guide 2021.
- [7] BS.2076 : Audio Definition Model <https://www.itu.int/rec/R-REC-BS.2076-2-201910-I/en> Retrieved 25. August 2021
- [8] Bleidt, R.L. (2015), "Installing the MPEG-H Audio Alliance's New Interactive and Immersive TV Audio System in Professional Production and Distribution Facilities", https://www.iis.fraunhofer.de/content/dam/iis/de/doc/ame/wp/FraunhoferIIS_TechnicalBrief_Installing-MPEG-H.pdf Retrieved 11. August 2021
- [9] "BS.2125: A serial representation of the Audio Definition Model", <https://www.itu.int/rec/R-REC-BS.2125-0-201901-I> Retrieved 6. August 2021.
- [10] "Overview of SMPTE ST 2110 Suite of Standards", <https://www.smpste.org/smpste-st-2110-faq> Retrieved August 2021
- [11] "Pilotprojekt Dialog+ - Mehr Sprachverständlichkeit im Fernsehen" <https://www.br.de/unternehmen/inhalt/technik/dialog-plus-tonspur-sprachverstaendlichkeit-fernseher-100.html> Retrieved 4. August 2021.
- [12] M. Torcoli et al. (2021), "Dialog+ in Broadcasting: First Field Tests Using Deep-Learning-Based Dialogue Enhancement." To appear, Int. Broadcasting Conv. (IBC) Technical Papers, 2021
- [13] "DaVinci Resolve Studio", www.blackmagicdesign.com/uk/products/davinciresolve/studio Retrieved 8. August 2021.
- [14] "MPEG-H Authoring Suite", <https://www.iis.fraunhofer.de/en/ff/amm/dl/software/mas.html> Retrieved 8. August 2021.
- [15] "Fraunhofer ADM Info Tool", <https://www.iis.fraunhofer.de/en/ff/amm/dl/software/adm-info-tool.html> Retrieved 8. August 2021
- [16] Spatial Audio Designer Plug-ins", <https://newaudiotechnology.com/products/spatial-audio-designer/> Retrieved 13. August 2021
- [17] "Auro-3D Creative Suite v3.0", <https://buy.auro-3d.com> Retrieved 13. August 2021
- [18] "Spatial Audio on demand: Music Streaming Services at a Glance", <https://www.vrtonung.de/en/music-streaming-spatial-audio/> Retrieved 9. August 2021.
- [19] Rieger, D. (2020), "Objektbasierte Musikproduktion", Masterthesis, Hochschule der Medien Stuttgart.
- [20] „360 Reality Audio Creative Suite“, <https://360ra.com> Retrieved 12. August 2021
- [21] Grewe, Y. et al. (2021), "Immersives & personalisiertes Audio für Rundfunk, Streaming &

Musik”. Webinar
<https://tonmeister.org/de/termine/liste/2021-06-08-mpeg-h-online-seminar-in-zusammenarbeit-mit-dem-fraunhofer-iis> Retrieved on 23. August 2021

[22] “Minnetonka AudioTools Server Adds MPEG-H Audio Authoring Workflow”,
<https://blogs.telosalliance.com/mpeg-h-audio-authoring-workflow-ats> Retrieved 9. August 2021.

[23] Poers, P. (2015), “Monitoring and Authoring of 3D Immersive Next Generation Audio formats”, in Proc. AES 139th Convention, New York City.

[24] “Spatial Audio Designer–Processor”,
<https://newaudiotechnology.com/products/spatial-audio-designer-processor/> Retrieved 9. August 2021.

[25] “A Live Mixing Console for Object-based Audio”,
<https://www.nhk.or.jp/str/english/news/2021/3.html>
Retrieved 9. August 2021.

[26] “3D Audio Market Size, Share & Industry Analysis [...]”,
<https://www.fortunebusinessinsights.com/3d-audio-market-102085> Retrieved 9. August 2021

[27] Grewe, Y. et al. (2021), “MPEG-H Audio Production Workflows for a Next Generation Audio Experience in Broadcast, Streaming, and Music”. To appear, Proc. AES 151st Convention, Las Vegas.